Resonant inelastic x-ray scattering in metals: A diagrammatic approach

A. M. Tsvelik,[1] R. M. Konik,[1] N. V. Prokof'ev,[2] and I. S. Tupitsyn[2]

[1]*Condensed Matter Physics and Materials Science Division, Brookhaven National Laboratory, Upton, New York 11973-5000, USA*
[2]*Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003, USA*

We develop a formalism to study the resonant inelastic x-ray scattering (RIXS) response in metals based on the diagrammatic expansion for its cross section. The standard approach to the solution of the RIXS problem relies on two key approximations: short-range potentials and noninteracting conduction electrons. However, these approximations are inaccurate for charged particles in metals, where the long-range Coulomb interaction and dynamic screening effects are very important. In this work, we study how to extract important information about collective excitations in the Coulomb plasma, plasmons, and electron-hole pairs from RIXS data. We find that single- and multi-plasmon excitations can easily be distinguished by positions of the corresponding peaks, singularities, and their intensities. We also discuss the hybrid processes, where plasmon emission is accompanied by excitation of electron-hole pairs and study how they manifest themselves.

I. INTRODUCTION

The resonant inelastic x-ray scattering (RIXS) technique holds tremendous promise for condensed matter physics [1–4]. It possesses several unique features, and its ability to reach high energy and momentum transfer enables studies of a wide range of collective excitations (see, for example, Ref. [5]). However, the RIXS cross section is not proportional to the excitation spectral function, and thus extracting the necessary information from data requires proper understanding of the underlying processes. After an incident photon creates a deep core hole, its strong potential disturbs the system, which results in the emission of multiple excitations and, correspondingly, leads to various nonlinear effects. Therefore, to distill properties of excitations, one needs to know how to separate single emission processes from the rest.

There exists a number of theoretical schemes to study the RIXS response [6–9]. In this work, we take a different track and develop a field-theoretical approach based on a diagrammatic expansion of the RIXS cross section (see also Refs. [10–12]). The advantage of this framework is in its universality—it allows one to address virtually any question about the system's behavior. Since RIXS is a second-order process (absorption followed by emission), we need to deal with the four-point time-ordered correlation function $\chi_R(\omega_i, \mathbf{q}_i, \omega_f, \mathbf{q}_f)$ for the dipole operators (see, for instance, Ref. [7]), where i and f stand for the initial (incoming) and final (outgoing) photons. Such a formulation is similar to the one used to study the Raman scattering process [13].

The overwhelming majority of theoretical efforts on the RIXS problem are based on two approximations: (i) a static short-range or contact core-hole potential and (ii) a noninteracting Fermi sea. However, both approximations oversimplify the nature of long-range Coulomb interactions and dynamic screening effects in metals. In the absence of dynamic screening, they lead to divergences that can be eliminated only by considering Fermi sea electrons as interacting via Coulomb forces as well. As far as we know, the RIXS problem in metals has never been systematically addressed beyond the formulation based on the above two approximations with the notable exception of Ref. [9] which incorporated Coulomb interactions into the dielectric function to study single-excitation emission process in layered copper-oxide systems.

The main goal of this work is to develop a more accurate understanding of various processes based on the Coulomb interactions, including emission of multiple excitations, and establish the framework for a high order diagrammatic expansion to the RIXS response (see, e.g., Ref. [14]).

There are two RIXS scenarios. In the first one, termed *indirect* RIXS, a deep core s electron is excited to a high-energy, potentially mobile, p state. The localized s hole possesses a strong potential generating low-energy collective excitations in the d band (see Ref. [6]). In this case, we have an s-p dipole, emitting d excitations during its lifetime, Γ^{-1}. The p electron eventually repopulates the s hole through a photon emission, leaving the d excitations behind. In the *direct* RIXS scenario, an electron from the s band is excited into the d band. Together, the d electron and s hole create collective excitations. During the hole's lifetime, the excited d electron moves away from the hole, and the photon is emitted when an electron from the occupied states recombines with the hole. In what follows, we focus on the indirect RIXS process (direct RIXS is briefly commented in the conclusions).

We work at $T = 0$ where energy transfer to the system is always positive. Typically, the core-hole life time is very short—of the order of a few femtoseconds [5]. This allows

one to limit the diagrammatic expansion for the RIXS cross section to just a few collective excitations. Below we take advantage of short hole's lifetime (SHLT) and describe the dynamic screening within the random phase approximation (RPA). For perturbative values of the Coulomb parameter r_s, this approach becomes exact. We concentrate on studying charge fluctuations (plasmons and particle-hole excitations) and discuss how signals from these collective excitations can be extracted from RIXS measurements by quantifying the contribution from the continuum of multiple excitations.

II. THE ANDERSON MODEL FOR INDIRECT RIXS

Our diagrammatic expansion for the RIXS cross section follows the standard scheme (see Ref. [13]), which can be illustrated by considering the Anderson model for core holes. To compute the cross section, one introduces two species of holes (labeled by $a = 1, 2$) localized at different space points ("sites") at a distance $\mathbf{R}_{12}$ from each other, and two species of p electrons. Then, the Anderson model can be formulated as follows:

$$H = H_s + H_d + H_p + H_{dd} + H_{sp} + H_{sd} + H_{pd};$$

$$H_s = \sum_{\sigma, a=1,2} \epsilon_s \, s_{\sigma,a}^\dagger \, s_{\sigma,a} + H_{s,\Gamma};$$

$$H_p = \sum_{k,\sigma,a} \epsilon_p(\mathbf{k}) \, p_{\mathbf{k},\sigma,a}^\dagger \, p_{\mathbf{k},\sigma,a}; \qquad (1)$$

$$H_d = \sum_{k\sigma} \epsilon_d(\mathbf{k}) \, d_{\mathbf{k},\sigma}^\dagger \, d_{\mathbf{k},\sigma};$$

$$H_{sd} = -\int d\mathbf{r} \, n_d(\mathbf{r}) \times \left[\frac{e^2}{|\mathbf{R}_1 - \mathbf{r}|} n_{s,1} + \frac{e^2}{|\mathbf{R}_2 - \mathbf{r}|} n_{s,2} \right],$$

where $s_{\sigma,a}$, $p_{\sigma,a}$, d_σ are the annihilation operators for the s core, p, and d electrons (correspondingly, $s_{\sigma,a}$ creates the s hole), $\sigma = \pm$ is the spin index, ϵ_s (here ϵ_s is k independent), $\epsilon_p(\mathbf{k})$, $\epsilon_d(\mathbf{k})$ are the corresponding dispersion relations, and n_d, n_s are number densities. $H_{s,\Gamma}$ defines the s hole with a finite lifetime Γ^{-1}. The interaction Hamiltonians H_{sd}, H_{pd}, and H_{dd} have a similar structure based on the Coulomb potential, $V_\mathbf{r} = e^2/r$ or $V_\mathbf{Q} = 4\pi e^2/Q^2$ (for brevity, we present explicitly only H_{sd}). Formally, in the orbital representation, these interactions are different, but this difference is insignificant for the purposes of our work.

By integrating the d electrons out within the RPA, we arrive at the model where s holes and p electrons are coupled by the action

$$S = S_0 + \int d\tau_1 d\tau_2 d\mathbf{r}_1 d\mathbf{r}_2 \rho(\mathbf{r}_1, \tau_1) U(\mathbf{r}_{12}, \tau_{12}) \rho(\mathbf{r}_2, \tau_2),$$

$$\rho(\mathbf{r}) = \sum_a \left[\delta(\mathbf{r} - \mathbf{R}_a) n_{s,a} + n_{p,a}(\mathbf{r}) \right], \qquad (2)$$

where S_0 is the bare action for s holes and p electrons and U is the dynamic part of the screened Coulomb potential [see Fig. 1(b)]. In this formulation, the correlation function, χ_R, responsible for the RIXS cross section, can be written as

$$\chi_R(\mathbf{R}_{12}; t_1, t_2, t) = \langle D_1(-t_1/2) D_1^+(t_1/2)$$
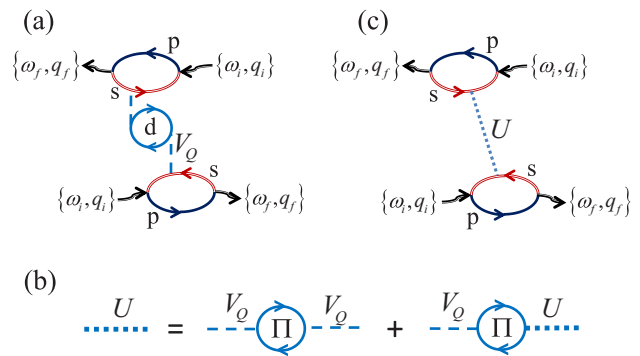$$\times D_2(t_2/2 + t) D_2^+(-t_2/2 + t) \rangle, \qquad (3)$$

FIG. 1. (a) The lowest-order diagram for the indirect RIXS cross section where the s hole excites charge density fluctuations in the d shell (the internal fermionic bubble) via the Coulomb potential V_Q (dashed lines). $(\omega_i, \mathbf{q}_i)/(\omega_f, \mathbf{q}_f)$ are the frequencies and momenta of incoming and outgoing photons. Energy and momentum transfer to the system are defined by $\Omega = \omega_i - \omega_f$ and $\mathbf{Q} = \mathbf{q}_i - \mathbf{q}_f$, respectively. (b) The Dyson equation for the screened dynamic interaction U obtained by summing up bubble diagrams based on V_Q and the polarization function, Π, of d electrons. (c) The lowest-order diagram for the indirect RIXS cross section in terms of U.

where $D_a^+ = p_{\sigma,a}^+ s_{\sigma,a}$ is the dipole creation operator on site $\mathbf{R}_a$. The RIXS cross section is extracted from the imaginary part of the analytically continued Fourier transform of this correlation function in direct analogy to the Raman scattering response [13]:

$$\chi_R(\omega_i, \mathbf{q}_i, \omega_f, \mathbf{q}_f) = \frac{1}{2\pi i} \lim_{\delta_1 > \delta_2 \to 0} [\tilde{\chi}(X_{1,i}; X_{1,f}; X_{2,f}; X_{2,i})$$
$$- \tilde{\chi}(X'_{1,i}; X'_{1,f}; X'_{2,f}; X'_{2,i})],$$
$$X_{1,i} = \{-\omega_i - i\delta_1, -\mathbf{q}_i\};$$
$$X'_{1,i} = \{-\omega_i - i\delta_2, -\mathbf{q}_i\};$$
$$X_{1,f} = \{\omega_f + i\delta_2, \mathbf{q}_f\};$$
$$X'_{1,f} = \{\omega_f + i\delta_1, \mathbf{q}_f\};$$
$$X_{2,f} = \{-\omega_f + i\delta_2, -\mathbf{q}_f\};$$
$$X'_{2,f} = \{-\omega_f + i\delta_1, -\mathbf{q}_f\};$$
$$X_{2,i} = \{\omega_i - i\delta_1, \mathbf{q}_i\};$$
$$X'_{2,i} = \{\omega_i - i\delta_2, \mathbf{q}_i\}. \qquad (4)$$

Self-explanatory notations for all variables are given in Figs. 1 and 2. In what follows, we expand χ_R into a diagrammatic series. The fully dressed Green's functions for both the s hole
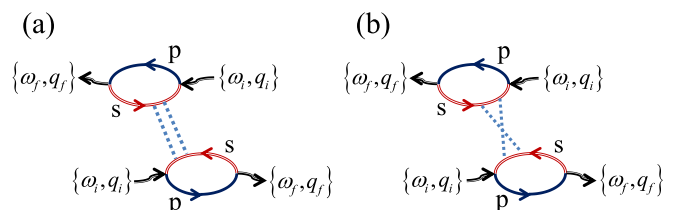

FIG. 2. Second-order diagrams for the RIXS cross section with two different topologies (a) and (b).

and the p electron (to account for interactions with d electrons) can be obtained within the diagrammatic Monte Carlo technique for polarons [14,15]. However, under the SHLT assumption, they can be approximated by their noninteracting expressions.

III. DIAGRAMMATIC REPRESENTATION FOR THE CROSS SECTION χ_R

If we focus on charge excitations for indirect RIXS, the relevant diagrammatic expansion is in the number of nonlocal interaction lines (2). The lowest order Feynman diagram for χ_R is shown in Fig. 1(a). The same diagram, but with a contact interaction V instead of the Coulomb potential V_Q, was considered in Ref. [10] for the case of quasi-one-dimensional (quasi-1d) Mott insulators. In contrast, Eq. (2) is written in terms of the dynamic interaction U based on the geometric series of bubble diagrams, as shown in Fig. 1(b). In the RPA, the polarization function Π is obtained from the product of two bare Green's functions for d electrons. While certainly insufficiently accurate for, say, transition metals, the RPA (exact in the limit of small r_s) does capture all qualitative features of the screening phenomenon. Thus, for Coulomb systems, the expansion order is defined by the number of the U lines.

In the SHLT limit, it is enough to keep only a few terms in the expansion. Two second-order diagrams are presented in Fig. 2. Calculation of the vertex functions, γ_v, to leading order in $1/\Gamma$ is presented in the Appendix. The contribution from the s-p loops does not depend on energy transfer Ω, but its momentum dependence is important. Since the d electrons interact with a neutral s-p excitation, all vertex functions vanish at zero momentum transfer. In particular, the vertex function squared for one $U(Q)$ line, Fig. 1(c), contains a factor of Q^4, while for the sum of two diagrams with $U(q_1)$ and $U(q_2)$ lines shown in Fig. 2, we have $|\gamma_v^{(2)}|^2 \propto (\mathbf{q_1 q_2})^2$. This difference leads to the suppression of contribution from the coherent plasmon at small momenta. The vertex function for the nth-order diagram is also proportional to $\Gamma^{-(n+1)}$.

The general structure of diagrams implies that emission of varying number of gapped excitations leads to different thresholds. This, in principle, allows one to distinguish the processes accompanied by additional excitations, provided that the spectral gaps are not too small compared to the bandwidths of these excitations. Given that contributions from higher order diagrams are suppressed by a factor $\propto \Gamma^{-2(n-1)}$, the most important corrections are determined by the second-order diagrams; see Fig. 2. For this reason, the main focus of this work is on the second-order processes. Note that the second-order contribution may dominate in the final answer if the first-order process has zero intensity in some frequency range or at small momentum transfer.

IV. PLASMON AND ELECTRON-HOLE EXCITATIONS

In this section, we explore how important information about charged excitations in metals can be extracted from the RIXS data by comparing first-order and second-order processes. In the SHLT limit, the imaginary part of the correlation
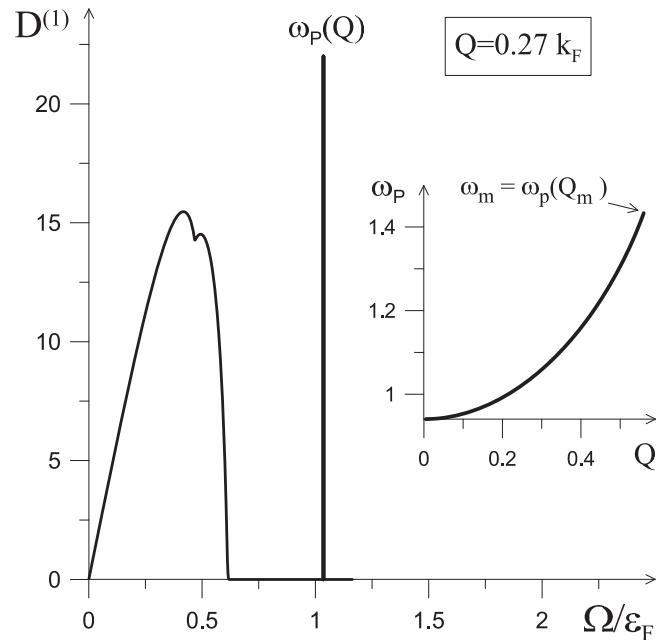

FIG. 3. First-order, see Fig. 1(c), contribution to the RIXS intensity (without the $|\gamma_v^{(1)}|^2$ factor) at $r_s = 1$. The plasmon dispersion relation is shown in the inset.

function responsible for the RIXS signal originates from the U functions, not from the vertex functions. Then, the first-order contribution can be written as

$$\chi_R^{(1)} = \left|\gamma_v^{(1)}(\omega_i, \omega_f; \mathbf{Q})\right|^2 D^{(1)}(\Omega, Q), \quad (5)$$

where $D^{(1)} = U''$ is the imaginary part of the screened potential

$$U(\Omega, Q) = \frac{4\pi e^2}{Q^2 - 4\pi e^2 \Pi(\Omega, Q)} - V_Q. \quad (6)$$

At zero temperature, the energy transfer to the system is non-negative, and in what follows we will implicitly assume that $\Omega \geqslant 0$. By approximating the d-band dispersion relation with the spherically symmetric expression $\epsilon_d = k^2/2m$, we easily obtain the result for $D^{(1)}(\Omega, Q)$ since $\Pi(\Omega, Q)$ in this situation reduces to the Lindhard function [16] (a typical curve is shown in Fig. 3). We work with units such that the Fermi momentum, $k_F = 1$, and Fermi energy, $\epsilon_F = 1$.

The main features of the $D^{(1)}$ spectrum are the sharp plasmon peak and the electron-hole continuum. Since plasmons are gapped excitations (their dispersion relation, $\omega_p(Q)$, is shown in the inset of Fig. 3), their contribution to intensity is separated from the low-energy particle-hole continuum at sufficiently small momenta; see Fig. 3. This separation becomes less pronounced at large momentum transfer and ultimately the plasmon peak merges with the continuum at

$$\Omega_m = \omega_p(Q_m) = v_F Q_m + \frac{Q_m^2}{2m}. \quad (7)$$

To recover the same basic properties, we radically simplify the Lindhard function while preserving exactly the key features of the important $Q \ll \kappa$ limit in terms of the Fermi liquid parameters. The corresponding approximation combines the

plasmon pole approximation [17,18] with Landau damping:

$$\Pi \approx -\rho_F + \frac{\rho_F \Omega}{2 v_F Q}\left[\ln\left|\frac{\Omega + v_F Q}{\Omega - v_F Q}\right| - i\pi\theta(v_F Q - \Omega)\right]. \quad (8)$$

Furthermore, in the same limit, the imaginary part of $U(\Omega, Q)$ can be separated into two distinct contributions, $U'' = D_{p\text{-}h}^{(1)} + D_{pl}^{(1)}$, associated with excitation of low-energy particle-hole pairs and gapped plasmon modes, respectively:

$$D_{p\text{-}h}^{(1)} = \frac{4\pi e^2}{\kappa^2}\frac{\pi\Omega}{2 v_F Q}\,\theta(v_F Q - \Omega), \quad (9)$$

$$D_{pl}^{(1)} = \frac{4\pi e^2}{Q^2}\frac{\pi}{2}\omega_p(Q)\,\delta(\Omega - \omega_p(Q)). \quad (10)$$

Finally, by using

$$\omega_p(Q) = \Omega_{pl} + \xi Q^2; \quad \xi = \frac{3}{10}\frac{v_F^2}{\Omega_{pl}}, \quad (11)$$

in Eq. (10), we correctly capture the plasmon dispersion at low momenta. By developing this effective description, we are now in position to address the problem of emission of multiple excitations in order to see whether and how their contributions can be separated from the first-order single-emission process.

In the $\Omega_{pl}/\Gamma \ll 1$ limit, the two diagrams shown in Fig. 2 give the same contribution. To account for the momentum dependence of the vertex function, we write it as $|\gamma_v^{(2)}(\mathbf{q}_1, \mathbf{q}_2)|^2 = (\gamma^2/\Gamma^2)\,f(\mathbf{q}_1, \mathbf{q}_2)$ with constant γ^2 and

$$f(\mathbf{q}_1, \mathbf{q}_2) = \frac{q_1^2\,q_2^2\,\mathcal{C}_{12}^2}{\left(q_1^2 + Q_d^2\right)\left(q_2^2 + Q_d^2\right)}, \quad (12)$$

where $\mathcal{C}_{12}$ is the cosine of the angle between vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, $Q_d \sim k_F$ is some high-momentum cutoff, and proceed with evaluating the integral

$$\int \frac{d\mathbf{q}\,d\omega}{(2\pi)^4} f(\mathbf{q}, \mathbf{Q} - \mathbf{q}) D^{(1)}(\omega, \mathbf{q}) D^{(1)}(\Omega - \omega, \mathbf{Q} - \mathbf{q}). \quad (13)$$

For two plasmons, the final result reads (without γ^2)

$$D_{pl}^{(2)} \propto \frac{\Omega_{pl}^2}{\Gamma^2}\frac{\pi e^4}{Q}\Delta(\Omega, Q), \quad (14)$$

with Δ given by

$$\Delta = \frac{1}{\Omega_{pl}^2}\int_0^{Q_m}\frac{q\,dq}{q^2 + Q_d^2}\frac{\omega_p(q)\,[\Omega - \omega_p(q)]\,\mathcal{C}_{12}^2}{[\Omega - \Omega_{pl} - \omega_p(q) + \xi Q_d^2]}$$
$$\times \Theta\left(1 - \left|\frac{2\omega_p(q) + \omega_p(Q) - \Omega_{pl} - \Omega}{2\xi Q q}\right|\right);$$
$$\mathcal{C}_{12}^2 = \frac{[\Omega - \Omega_{pl} - \omega_p(Q)]^2}{4[\omega_p(q) - \Omega_{pl}][\Omega - \Omega_{pl} - \omega_p(q)]}, \quad (15)$$

featuring singularities in derivatives distinctly related to the plasmon spectrum; see Fig. 4. For a dispersionless plasmon, $\Delta(\Omega)$ is proportional to $\delta(\Omega - 2\Omega_{pl})$. Thus, the one- and two-plasmon processes can be distinguished by positions and shapes of the corresponding peaks and intensities. While the one-plasmon process results in the sharp peak at $\omega_p(Q)$, the two-plasmon curve is broad and is shifted outside of the
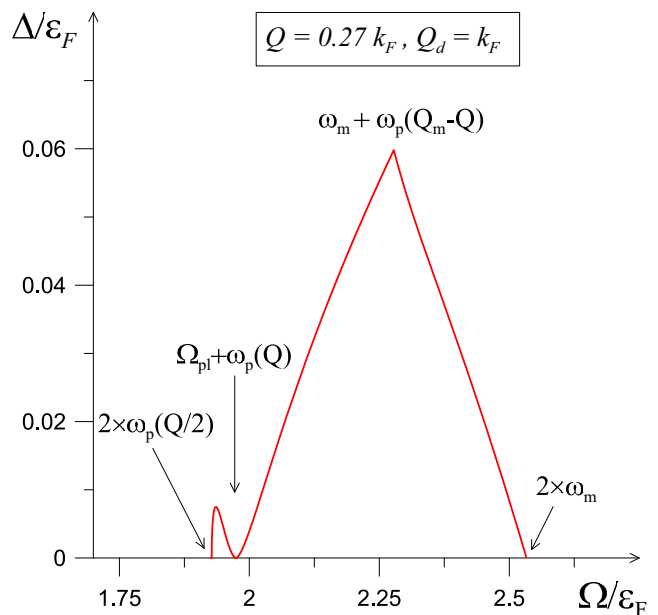

FIG. 4. Second-order contributions to the RIXS intensity: Δ as a function of frequency Ω at $r_s = 1$. [For the simplified plasmon dispersion (11), $Q_m \approx 0.5055 k_F$.]

plasmon dispersion relation. The ratio of intensities goes as

$$\frac{\gamma^2 D_{pl}^{(2)}}{|\gamma^{(1)}(Q)|^2 D_{pl}^{(1)}} \propto \frac{\Omega_{pl} e^2 k_F^3}{\Gamma^2 Q^2}. \quad (16)$$

At low momenta, the two-plasmon process will produce a stronger signal than the single-plasmon one.

There is also a second-order hybrid process, when plasmon emission is accompanied by an excitation of a particle-hole pair. The corresponding spectrum overlaps with the single-plasmon peak,

$$D_{pl,p\text{-}h}^{(2)} \propto \frac{\Omega_{pl}^2}{\Gamma^2}\frac{\pi^2 e^2}{4 k_F^2} F(z, y), \quad (17)$$

where $z = (\Omega - \Omega_{pl})/\Omega_{pl}$, $y = v_F Q/\Omega_{pl}$, and

$$F(z, y) = \int_0^{x_m} dx \frac{(1 + x)(z - x)}{x\left(x + \xi Q_d^2/\Omega_{pl}\right)}$$
$$\times [Y(x, y, t_2) - Y(x, y, t_1)], \quad (18)$$

with $x_m = \min[z, (\Omega_m - \Omega_{pl})/\Omega_{pl}]$ and

$$Y = \frac{u}{4}\left[\frac{t}{3}(6a - 3b + t^2) + \frac{(a - b)^2 \tan^{-1}[t/\sqrt{b}]}{\sqrt{b}}\right];$$
$$u = (3/10)y^2; \quad a = \frac{x - u}{u}; \quad b = \frac{Q_d^2}{Q^2}. \quad (19)$$

The dependence on Y is through the restrictions on the domain of integration:

$$t_2 = 1 + \sqrt{x/u};$$
$$t_1 = \max(|1 - \sqrt{x/u}|, (z - x)/y) < t_2. \quad (20)$$

For small z and y and $z \gg y^2$, we have $F \sim z^3$; if $z \ll y^2$, then $F \sim z^{5/2}$. The entire functional dependence in shown in Fig. 5
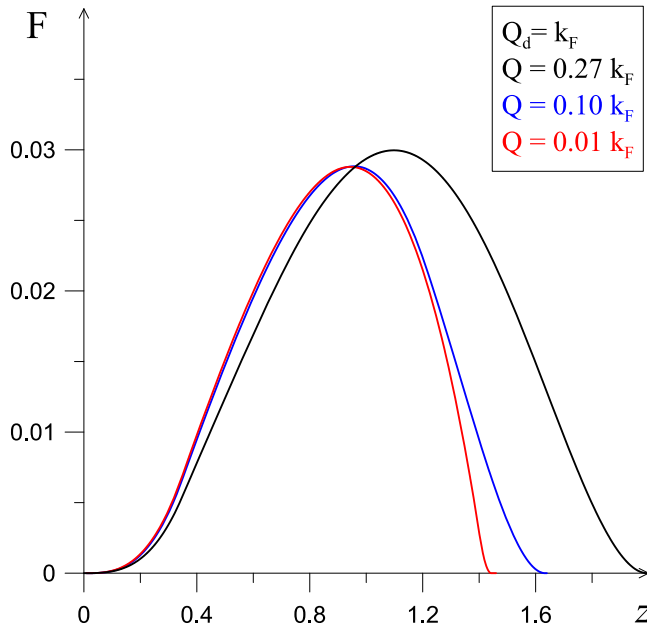
FIG. 5. Second-order contributions to the RIXS intensity: The hybrid amplitude F as a function of $(\Omega - \Omega_{pl})/\Omega_{pl}$ for three values of external momentum transfer ($Q = 0.27k_F$, black; $Q = 0.1k_F$, blue; and $Q = 0.01k_F$, red) at $r_s = 1$.

for several values of Q and $Q_d = k_F$. For large values of Ω, the simplified description, Eqs. (9) and (10), suitable for analytic treatment of long-wave excitations, loses its accuracy because momenta of two excitations may compensate each other. In this case, Eq. (8) needs to be replaced with the exact Lindhard function [16].

At small momentum Q, this contribution can also be stronger in comparison to the single-plasmon one. On the one hand, since there is no threshold for emission of particle-hole pairs, the hybrid process can contaminate the single-plasmon measurement. On the other hand, the intensity of this process involves powers of $(\Omega - \omega_p)$ and thus fades in the vicinity of the single-plasmon threshold at Ω_{pl}. Away from the threshold, this contribution is very broad and can be easily discriminated from sharp peaks.

V. CONCLUSIONS

We have addressed the problem of indirect RIXS in metals and provided a quantitative framework for understanding the key features of the RIXS signal in Coulomb systems, including two-excitation processes. We have done so based on a field theoretic approach that takes into account long-range Coulomb interactions. As a test bed, we have studied the Coulomb gas. Here we have been able to distinguish single-particle from multiparticle excitations. We have found that the two-plasmon and hybrid signals can be stronger in intensity than the single-excitation spectra at small momentum transfer and that they have different distinctive characteristics.

We have used here an RPA approach, valid for small values of r_s (including $r_s = 1$). For larger values of r_s, one has to consider diagrams accounting for self-energy, polarization, and vertex function corrections, as well as diagrams that do not factor into the product of U lines. We plan to implement this program within diagrammatic Monte Carlo method in an approximation-free way; see, e.g., Ref. [19].

It is worth mentioning that the same technique can be applied to study the direct RIXS process. The short hole's lifetime approximation can be removed by computing fully dressed Green's functions obtained in the way described in Ref. [14]. One will also need to account for the interband transitions and distinguish contributions originating from localized holes and mobile p, d electrons.

APPENDIX

We introduce the notations for our computations below:

$$G^{(p)}(\omega, k) = [i\omega_n - \epsilon(k)]^{-1},$$
$$G^{(s)}(\omega) = [i\omega_n + i\Gamma \mathrm{sgn}\omega_n]^{-1}. \quad \text{(A1)}$$

Now there are two vertex diagrams with one dashed line. The simplest diagram with one Coulomb dashed line consists of two parts depending on whether this line is connected to the hole:

$$\gamma^{(h)} = 2T \sum_{\omega} \int \frac{d^3k}{(2\pi)^3} G^{(s)}(\omega) G^{(s)}(\omega + \omega_i - \omega_f) G^{(p)}(\omega + \omega_i, k)$$

$$= 2T \sum_{n} \int \frac{d^3k}{(2\pi)^3} \frac{1}{i(\omega_n + \omega_i) - \epsilon_k} \int \frac{dx\rho(x)}{i\omega_n - x} \int \frac{dy\rho(y)}{i(\omega_n + \omega_i - \omega_f) - y}$$

$$= 2i \int \frac{d^3k}{(2\pi)^3}\theta(\epsilon_k) \int \frac{dx\rho(x)}{-i\omega_i + \epsilon_k - x} \frac{dy\rho(y)}{-i\omega_f + \epsilon_k - y} \rightarrow 2 \int \frac{\nu_p(\epsilon)d\epsilon}{(\omega_i + \epsilon_s - \epsilon + i\Gamma)(\omega_f + \epsilon_s - \epsilon + i\Gamma)}, \quad \text{(A2)}$$

where the factor of 2 comes from the summation over spin and

$$\rho(x) = \frac{1}{\pi}\frac{\Gamma}{\Gamma^2 + (x - \epsilon_s)^2},$$

or to the p electron:

$$\gamma^{(e)} = 2T \sum_{\omega} \int \frac{d^3k}{(2\pi)^3} G^{(s)}(\omega) G^{(p)}(\omega + \omega_i - \omega_f, k) G^{(p)}(\omega + \omega_f, k - q)$$

$$= 2T \sum_{n} \int \frac{d^3k}{(2\pi)^3} \frac{1}{[i(\omega_n + \omega_i) - \epsilon_k][i(\omega_n + \omega_f) - \epsilon_{k-q}]} \int \frac{dx \rho(x)}{i\omega_n - x}$$

$$= \rightarrow 2 \int \frac{d^3k}{(2\pi)^3} \frac{1}{\omega_i - \omega_f + i\delta - \epsilon_{k-q} + \epsilon_k} \left[\frac{\theta(\epsilon_k)}{\omega_f + i\Gamma - \epsilon_k + \epsilon_s} - \frac{\theta(\epsilon_{k-q})}{\omega_i + i\Gamma - \epsilon_{k-q} + \epsilon_s} \right]. \tag{A3}$$

Here the arrows denote the analytic continuation described in the main text.

Since ϵ_k is always positive, we have

$$\gamma^{(e)} = 2 \int \frac{d^3k}{(2\pi)^3} \frac{1}{(\omega_f + i\Gamma - \epsilon_k + \epsilon_s)(\omega_i + i\Gamma - \epsilon_{k-q} + \epsilon_s)}, \tag{A4}$$

so that the entire vertex becomes

$$\gamma^{(1)} \equiv \gamma^{(h)} - \gamma^{(e)} = 2 \int \frac{d^3k}{(2\pi)^3} \frac{\epsilon_k - \epsilon_{k-q}}{(\omega_f + i\Gamma - \epsilon_k + \epsilon_s)(\omega_i + i\Gamma - \epsilon_{k-q} + \epsilon_s)(\omega_i + i\Gamma - \epsilon_k + \epsilon_s)}. \tag{A5}$$

It is universal only at small momenta. Adopting $\epsilon_k = k^2/2m$, we obtain

$$\gamma^{(1)}(q) = \frac{q^2}{2m}(2m/\Gamma)^{3/2} \left[\frac{1}{\pi^2} \int_0^\infty \frac{dx x^{1/2}}{(i-x)^3} \right], \tag{A6}$$

(the part linear in q cancels in the integration over **k** due to the inversion symmetry). Otherwise, the vertex is model dependent.

The second-order vertex [with two wavy lines, one with (ω_1, q_1) and another with $(\Omega - \omega_1, Q - q_1)$] is (here we have included ϵ_s in the definition of ϵ_k)

$$\gamma^{(2)} = \gamma^{hh} + \gamma^{ee} - 2\gamma^{he} = 2 \int \frac{d^3k}{(2\pi)^3} [\Gamma_1 + \Gamma_2 + \Gamma_3];$$

$$\Gamma_1 = \frac{1}{[i\Gamma - \omega_f - \epsilon_{k-q_f}][i\Gamma - (\omega_f + \omega_1) - \epsilon_{k-q_f-q_1}][i\Gamma - \omega_i - \epsilon_{k-q_i}]};$$

$$\Gamma_2 = -\frac{2i\Gamma - (\omega_i + \omega_f) - \epsilon_{k-q_1-q_f} - \epsilon_{k-q_i}}{[i\Gamma - \omega_f - \epsilon_{k-q_f-q_1}][i\Gamma - (\omega_f + \omega_1) - \epsilon_{k-q_f-q_1}][i\Gamma - \omega_i - \epsilon_{k-q_i}][i\Gamma + \omega_1 - \omega_i - \epsilon_{k-q_i}]};$$

$$\Gamma_3 = \frac{1}{[i\Gamma - \omega_f - \epsilon_k][i\Gamma - (\omega_i - \omega_1) - \epsilon_k][i\Gamma - \omega_i - \epsilon_k]}. \tag{A7}$$

When the external momenta are zero, it also vanishes. It can be rewritten as

$$\gamma^{(2)} = 2 \int \frac{d^3k}{(2\pi)^3} \left\{ \frac{1}{[i\Gamma - (\omega_f + \omega_1) - \epsilon_k][i\Gamma - \omega_i - \epsilon_{k+Q-q_1}]} \left(\frac{1}{i\Gamma - \omega_f - \epsilon_{k+q_1}} - \frac{1}{i\Gamma - \omega_f - \epsilon_k} \right) \right.$$

$$\left. + \frac{1}{[i\Gamma + (-\omega_i + \omega_1) - \epsilon_k][i\Gamma - \omega_i - \epsilon_k]} \left(\frac{1}{i\Gamma - \omega_f - \epsilon_k} - \frac{1}{i\Gamma - \omega_f - \epsilon_{k-Q+q_1}} \right) \right\}. \tag{A8}$$

At small momenta, $\gamma^{(2)}(q_1, Q - q_1) \sim [q_1^2 - (Q + q_1)^2]$. However, this leading contribution cancels when one adds up the two diagrams in Figs. 2(a) and 2(b) with interchanged legs:

$$\gamma^{(2)}(q_1, Q - q_1) + \gamma^{(2)}(Q - q_1, q_1) \sim \int d^3k (\epsilon_k - \epsilon_{k+q_1})(\epsilon_k - \epsilon_{Q-q_1+k}). \tag{A9}$$

At small momenta, this equals $\sim [\mathbf{q}_1(\mathbf{Q} - \mathbf{q}_1)]$.

---

[1] M. Dean, J. Magn. Magn. Mater. **376**, 3 (2015).

[2] F. Vernay, B. Moritz, I. S. Elfimov, J. Geck, D. Hawthorn, T. P. Devereaux, and G. A. Sawatzky, Phys. Rev. B **77**, 104519 (2008).

[3] C.-C. Chen, B. Moritz, F. Vernay, J. N. Hancock, S. Johnston, C. J. Jia, G. Chabot-Couture, M. Greven, I. Elfimov, G. A.

Sawatzky, and T. P. Devereaux, Phys. Rev. Lett. **105**, 177401 (2010).

[4] M. Guarise, B. Dalla Piazza, H. Berger, E. Giannini, T. Schmitt, H. Ronnow, G. Sawatzky, J. van den Brink, D. Altenfeld, I. Eremin, and M. Grioni, Nat. Commun. **5**, 5760 (2014).

[5] L. J. P. Ament, M. van Veenendaal, T. P. Devereaux, J. P. Hill, and J. van den Brink, Rev. Mod. Phys. **83**, 705 (2011).

[6] R. S. Markiewicz, J. J. Rehr, and A. Bansil, Phys. Rev. Lett. **112**, 237401 (2014).

[7] M. Kanasz-Nagy, Y. Shi, I. Klich, and E. A. Demler, Phys. Rev. B **94**, 165127 (2016).

[8] Y. Shi, A. J. A. James, E. Demler, and I. Klich, Phys. Rev. B **96**, 155101 (2017).

[9] M. Hepting, L. Chaix, E. W. Huang, R. Fumagalli, Y. Y. Peng, B. Moritz, K. Kummer, N. B. Brookes, W. C. Lee, M. Hashimoto, T. Sarkar, J.-F. He, C. R. Rotundu, Y. S. Lee, R. L. Greene, L. Braicovich, G. Ghiringhelli, Z. X. Shen, T. P. Devereaux, and W. S. Lee, Nature (London) **563**, 374 (2018).

[10] T. Nomura and J. Igarashi, J. Phys. Soc. Jpn. **73**, 1677 (2004); Phys. Rev. B **71**, 035110 (2005).

[11] P. M. Platzman and E. D. Isaacs, Phys. Rev. B **57**, 11107 (1998).

[12] T. P. Devereaux, A. M. Shvaika, K. Wu, K. Wohlfeld, C. J. Jia, Y. Wang, B. Moritz, L. Chaix, W.-S. Lee, Z.-X. Shen, G. Ghiringhelli, and L. Braicovich, Phys. Rev. X **6**, 041019 (2016).

[13] A. M. Shvaika, O. Vorobyev, J. K. Freericks, and T. P. Devereaux, Phys. Rev. B **71**, 045120 (2005).

[14] I. S. Tupitsyn and N. V. Prokof'ev, Phys. Rev. B **99**, 245122 (2019).

[15] N. V. Prokof'ev and B. V. Svistunov, Phys. Rev. Lett. **81**, 2514 (1998).

[16] J. Lindhard, Mat. Fys. Medd. K. Dan. Vidensk. Selsk. **28**, 1 (1954).

[17] B. I. Lundqvist, Phys. Kondens. Mater. **6**, 193 (1967); **6**, 206 (1967).

[18] L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, New York, 1969), Vol. 23, pp. 1–181.

[19] I. S. Tupitsyn, A. S. Mishchenko, N. Nagaosa, and N. Prokof'ev, Phys. Rev. B **94**, 155145 (2016).